DOCUMENT RESUME

ED 179 090 FL 010 658

AUTHOR Crain, Stephen; Ocker, Pamela L.
TITLE A Semantic Constraint on Syntactic Parsing.
PUB DATE Dec 78
NOTE 10p.; Paper presented at the Annual Meeting of the Linguistic Society of America (Boston, MA, December 28-30, 1978)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Deep Structure; Language Processing; Language Research; *Linguistic Theory; Phrase Structure; *Semantics; *Sentence Structure; Surface Structure; *Syntax
IDENTIFIERS *Parsing; *Semantic Interference Effect

ABSTRACT

This research examines how semantic information influences syntactic parsing decisions during sentence processing. In the first experiment, subjects were presented lexical strings having syntactically identical surface structures but with two possible underlying structures: "The children taught by the Berlitz method," and "The teachers taught by the Berlitz method." Strings of the first type were processed as complex noun phrases, since the initial noun (children) was semantically more likely to be interpreted as the object of the following verb (taught). Strings of the second type were processed as complete sentences, following the canonical-sentoid strategy, since the initial noun (teachers) was a more likely subject of the verb. In a second experiment, strings of both types were followed by verb phrases: "The children taught by the Berlitz method passed the test," and "The teachers taught by the Berlitz method passed the test." Sentences like the first were parsed without difficulty as complete sentences. However, sentences like the second produced "garden path" effects. Since these results can be attributed to the semantic determination of subjects and objects, syntactic or "perceptual" strategies do not appear to be sufficient to account for the formulation of underlying syntactic structures. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED179090

A Semantic Constraint on Syntactic Parsing

Stephen Crain

University of California, Irvine

Pamela L. Coker

Claremont Graduate School

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Stephen Crain

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FL010658

Paper presented at the Linguistic Society of America Annual Meeting, December 28-30, 1978

© Stephen Crain & Pamela L. Coker

A Semantic Constraint on Syntactic Parsing

This research is part of an ongoing effort to investigate the interaction between syntax and semantics during sentence processing. The experiments reported here provide empirical support for an interactive model of sentence processing in which syntactic parsing and semantic composition interact on-line to assign both syntactic and semantic analyses. This model of sentence processing was originally formulated by Kimball (1973, 1975) and has been given some previous empirical support by Marslen-Wilson (1975) and Tannenhaus & Seidenberg (1977). It is an alternative to the purely syntactic accounts of sentence processing recently proposed by Marcus (1977) and Frazier & Fodor (1978). It also differs from an account utilizing cognitive perceptual strategies, like the Canonical Sentoid Strategy proposed by Fodor, Bever & Garrett (1974). This paper reports on a series of experiments which show how lexical semantic information can influence syntactic parsing.

As a vehicle to show the interaction of syntactic and semantic processing, garden path sentences were used. Garden path sentences serve as a test case for the adequacy of any theory of sentence processing. Classic examples discussed in the literature include the following:

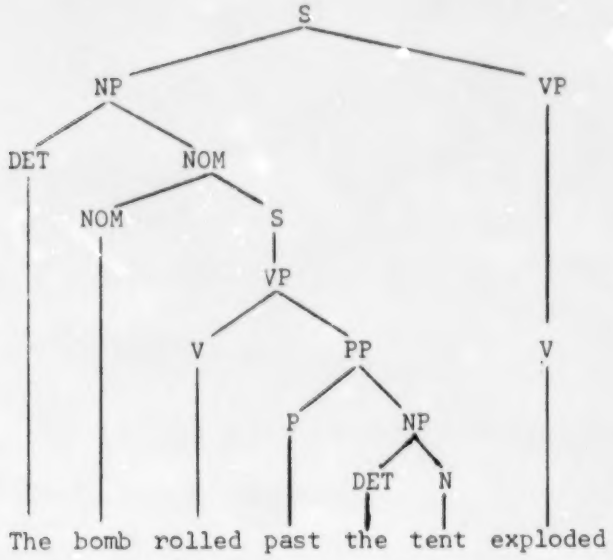
1. The bomb rolled past the tent exploded.
2. The pitcher tossed the ball tossed the ball.
3. The horse raced past the barn fell.

Bever (1974) as well as Chomsky & Lasnik (1977) have argued convincingly that the unacceptability of garden path sentences is due to processing difficulties and not due to their grammatical status.

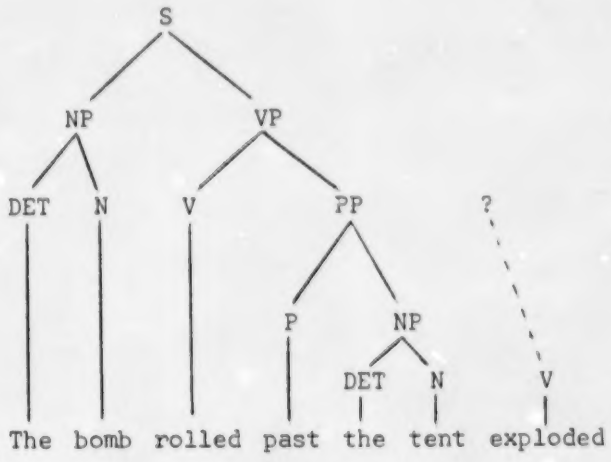
Garden path sentences result from the omission of all syntactic markers which signal that one is parsing a Complex NP, ie. an NP with a relative clause.

As the two tree diagrams below show, there are two well-formed syntactic analyses up to the point of the final verb phrase. It is here that the processing difficulty arises if the parser has analyzed the ambiguous sub-string as a sentence.

(a) Grammatical Surface Structure



(b) Parsed Structure (in Garden Paths)



As can be seen from their unambiguous counterparts, garden path sentences become acceptable when there is either a relative pronoun (as in 1' below), or when the past participial form of the verb is distinct from its simple past form (as in 2'). 3' shows that a prior syntactic set can also aid the parser in generating the grammatically correct structure. This is just a partial list of the syntactic cues used to signal a Complex NP in English.

1'. The bomb that was rolled past the tent exploded.

2'. The pitcher thrown the ball tossed the ball.

3'. The horse ridden by the water tower won the race.
But the horse raced past the barn fell.

Previous research has assumed that when all the syntactic cues to the correct structure are omitted, garden path sentences invariably result. This assumption of invariance has led psycholinguists to postulate models of sentence processing which predict that all sentences omitting these cues will lead the parser down the garden path. Only an interactive model predicts that lexical semantic information can be used to direct the parser in choosing what turns out to be the correct grammatical analysis. Thus, it does not predict that all sentences which omit these syntactic cues will result in garden path sentences.

The classic examples of garden path sentences are unacceptable not simply because they omit syntactic cues to their "correct" structure. It is also more plausible semantically to parse their initial lexical items as sentences rather than Complex NPs. If this is correct, one should be able to construct sentences where lexical information semantically favors the alternative, Complex NP analysis. Sentences produced in this way would be both grammatical and acceptable.

4.) The {tenant / postman} delivered junk mail threw it in the trash.

5.) The {cheater / genius} furnished the answers passed the test.

6.) The {performer / florist} sent the flowers was thrilled.

Examples 4-6 are sentence pairs which omit the relevant syntactic cues but where one member is semantically more plausible when the initial lexical items are parsed as Complex NPs. The other member of each pair differs only by the substitution of a single lexical item, but it is more likely semantically to lead the parser down the garden path. Take the first example (4). An interactive model predicts that the parser is more likely to find acceptable "The tenant delivered junk mail threw it in the trash", as opposed to "The postman delivered junk mail threw it in the trash." Similarly, in sentence pairs (5) and (6), it is predicted that "cheater" and "performer" will not lead the parser down the garden path but "genius" and "florist" will.

What is predicted is that the meaning of the initial noun, when combined with the meaning of the following verb-phrase, can serve to control the parser's choice between the two possible syntactic analyses of the sub-string. Consequently, the entire sentence is parsed either on a garden path or off of one depending on semantic information. To test this hypothesis, 30 sentence pairs (Exper. 1, pg. 5) were given to subjects to see if these sentence pairs are indeed parsed differently. These target sentences were hidden among 90 unambiguous sentences. Subjects heard all the sentences presented in a monotone, each subject receiving only one member of any target-sentence pair. Following each sentence, the subjects performed a simple reaction time task. Immediately following this task, they heard a statement which was based on the previous sentence. Subjects were asked to say whether this statement was True or False of the sentence. For example, in the first experiment they heard the target string "The students instructed about the assignment." Then they heard the statement "The assignment was given by the students" and were asked to judge the truth value of this statement. If subjects said False to this, it was inferred that they had parsed the target string as a Complex NP. If they said True, this judgement was interpreted as indicating that they had parsed the string as a complete sentence.

In this first experiment, target strings were structurally ambiguous between a complete sentence and a Complex NP (and hence, a fragment of a sentence). It was predicted that "The students instructed about the assignment" would be parsed as a Complex NP while "The professors instructed about the assignment" would be parsed as a complete sentence. Subjects were <u>not</u> told to expect sentence fragments, and only 15 out of 120 sentences they encountered could be plausibly interpreted as sentence fragments. Yet, even with this kind of syntactic set to take the strings as <u>sentences</u>, the subjects parsed the predicted Complex NPs as sentence fragments 43% of the time. Subjects parsed the strings predicted to be sentences as complete sentences 84% of the time. These results are consistent with an interactive model which allows semantic information derived from lexical items to control syntactic parsing.

Experiment 1:

Target String (fragment)

The students instructed about the assignment
The professors instructed about the assignment

Judgement

The assignment was given by the students.
The assignment was given by the professors.

Results

Predicted Response	Percentage of Predicted Responses	Syntactic Parse
(False)	43%	Complex NP
(True)	84%	Sentence

These results were extended to garden path sentences in a second experiment. For this experiment, the 60 target strings used in the first experiment were filled out by adding verb phrases. Here we anticipated that those strings previously parsed as Complex NPs would be correctly parsed as grammatical sentences when filled-out. We predicted garden path effects only with those target strings which were parsed as sentences in the previous experiment. This prediction was confirmed. In sentences like "The students instructed about the assignment were confused", subjects assigned the correct grammatical structure 79% of the time. On the other hand, subjects only assigned the correct grammatical structure 54% of the time in sentences like "The professors instructed about the assignment were confused." So almost half of the time subjects were led down the garden path for target strings which were parsed as complete sentences in our previous experiment.

Experiment 2:

Target String (Garden Path)

The students instructed about the assignment were confused.
The professors instructed about the assignment were confused.

Judgement

The assignment was given by the students.
The assignment was given by the professors.

Results

Correct Response	Percentage of Correct Responses	Syntactic Parse
(False)	79%	Sentence
(False)	54%	Sentence

Subjects were told to make their judgement of the truth value of the statement based on the previous string they encountered. One can infer from this truth value judgement which of the two possible syntactic analyses they assigned to each target string. It is conceivable, however, that subjects were simply responding True or False on the basis of the semantic plausibility of the judgement statement alone. For instance, they might have responded False to the judgement "The assignment was given by the students" not because it was a False statement based on how they parsed the target string, but because it is semantically implausible that students give assignments.

A control condition was added to ensure that the judgements of truth value made were actually based on the syntactic parsing of the target string. A set of semantically odd but unambiguous sentences was constructed; e.g. "The mother (that was) disciplined by the child learned a lesson." These sentences were again

followed by sentences requiring a judgement, e.g. "The child misbehaved." For semantically odd sentences like this one, subjects made the correct truth value judgements 84% of the time. If subjects had been basing their judgements purely on semantic plausibility, they would not have any correct inferences in this control condition. It is assumed then that the subjects in the first two experiments did base their truth value judgements on the target strings. Our results show that people are led down the garden path primarily when lexical semantic constraints do not suggest the correct syntactic structure. In the second experiment, out of 60 potential garden path sentences, only 1/3 produced inferences consistent with a garden path analysis.

Theories which do not hypothesize that semantic information can influence syntactic parsing could not have anticipated these results. First, Fodor, Bever & Garrett hypothesized the Canonical Sentoid Strategy to account for the unacceptability of garden path sentences. This strategy is a structure-independent mapping from surface syntactic categories to the semantic interpretation of a sentence. This perceptual strategy always takes the initial noun in a clause to be the actor of the next inflected verb. Our results show that the initial noun is sometimes not the actor of the verb that follows. Furthermore, the fact that it plays a different semantic role actually influences the syntactic parsing of the sentence. A perceptual strategy account seems to preclude the result that when people parse sets of sentences like these they actually come up with two distinct syntactic analyses, one of which enables them to successfully parse many potential garden path sentences.

Both of the syntactic parsing models mentioned at the beginning of the paper also fail to account for our findings. The syntactic parser proposed by Marcus does not employ any semantic information in syntactic decision-making and Frazier & Fodor's syntactic parser utilizes semantic information only to reject anomolous readings of syntactically well-formed strings. Semantics in these theories either plays no role or plays too weak a role to actually control on-line syntactic decisions. Marcus and Frazier & Fodor have made important contributions to our understanding of syntactic parsing. However, we feel they have underestimated the role semantics actually has in sentence processing. We have demonstrated that there is an interaction between the syntactic and semantic components where lexical semantic constraints sometimes control syntactic parsing.

References

Bever, T.G. The psychology of language and structuralist investigations of nativism. In G. Harman (Ed.) On Noam Chomsky. New York: Anchor Press, 1974.

Chomsky, N., & Lasnik, H. Filters and control. Linguistic Inquiry, 1977, 8, 3, 425-504.

Fodor, J.A., Bever, T.G., & Garrett, M.F. The psychology of language: An introduction to psycholinguistics and generative grammar. New York: McGraw-Hill, 1974.

Frazier, L., & Fodor, J.E. The sausage machine: A new two-stage parsing model. Cognition, 1978, 2, 291-325.

Kimball, J. Seven principles of surface structure parsing in natural language. Cognition, 1973, 2, 15-47.

Kimball, J. Predictive analysis and over-the-top parsing. In J. Kimball (Ed.) Syntax and semantics, vol. 4. New York: Academic Press, 1975.

Marcus, M. A theory of syntactic recognition for natural language. Unpublished doctoral dissertation, M.I.T., 1977.

Marslen-Wilson, W.D. Sentence perception as an interactive parallel process. Science, 1975, 189, 226-228.

Tannenhaus, M.K., & Seidenberg, M. Psychological constraints on grammars: On trying to put the real back in "psychological reality". In Papers from the thirteenth regional meeting of the Chicago linguistic society. Chicago: Chicago Linguistic Society, 1977.